

### **REMARKS**

Applicants appreciate the Examiner's thorough review of the present application, and respectfully request reconsideration in light of the preceding amendments and the following remarks.

Claims 3-4, 13-22 and 25-26 have been cancelled without prejudice or disclaimer. The remaining claims have been revised to improve claim language. The amended claims find solid support in the original specification, e.g., paragraphs 0050, 0053-0055, 0059, 0065, 0067, 0069 and FIGs. 2A-2B and 3. No new matter has been introduced through the foregoing amendments.

1. Per the Examiner's request, Applicants submit herewith a *sworn* English translation of the second priority document, i.e., JP 2003-105867. All priority claims in the instant application have been perfected.

2. The Title has been revised to be more descriptive. However, if the revised Title is considered by the Office to be insufficiently descriptive, the Office is kindly asked to suggest an appropriate Title.

3. Withdrawal of the previously indicated allowable subject matter of claim 7 in view of U.S. Patent Application Publication No. 2003/0028558 in the name of *Kawatani* is noted. Applicants respectfully submit that *Kawatani* is not prior art to the present invention because it is not "by another,"<sup>1</sup> i.e., the reference invention was made by the same inventor of this application as evident from the attached Rule 132 Declaration.

Withdrawal of the rejections relying on *Kawatani* is believed appropriate and therefore respectfully requested.

---

<sup>1</sup> See MPEP, section 706.02(b) (Filing an affidavit or declaration under 37 CFR 1.132 showing that the reference invention is not by "another." See MPEP § 715.01(a), § 715.01(c), and § 716.10).

4. The Examiner's indication of allowable subject matter of claims 9-12 is noted with appreciation. Appropriate amendments have been made to place the claims in condition suitable for allowance.

5. The *35 U.S.C. 112, second paragraph* rejection of claims 1-31 is noted. Although Applicants do not necessarily agree with the Examiner's position that the recited feature renders the claims indefinite, amendments have nevertheless been made solely for the purpose of expediting prosecution, i.e., the allegedly indefinite feature has been removed/revised.

6. The *35 U.S.C. 112, second paragraph* rejections of claims 7, 9-12 are believed overcome in view of the above amendments.

7. The *35 U.S.C. 101* rejection of all claims is believed overcome in view of the amendments as discussed above with respect to section 5 of this Amendment.

8. The *35 U.S.C. 101* rejection of claims 17-22 is noted. Applicants respectfully disagree with the Examiner's position because the claimed computer program product is not directed to software *per se*. Rather, the claims are directed to hardware containing thereon a program as explicitly recited in the claim language, and hence, are directed to statutory subject matter. Solely for the purpose of expediting prosecution, the rejected claims have been canceled to simplify the issues.

9. The art rejections of all claims are noted. Although Applicants do not necessarily agree with the Examiner's position, amendments have nevertheless been made solely for the purpose of expediting prosecution. Applicants respectfully submit that the applied references singly or in combination fail to disclose, teach or suggest all limitations of, at least, the amended independent claims. For example, amended claim 1 is now directed to

1. A method of clustering documents each having one or plural document segments in an input document set, said method comprising:

(a) obtaining a co-occurrence matrix for each input document, and obtaining an input document frequency matrix for the set of input documents based on occurrence frequencies of terms or term pairs appearing in the set of input documents;

(b) selecting a seed document from a set of remaining documents that are not included in any cluster existing at that moment, and constructing a current cluster of an initial state based on the seed document, wherein said selecting and constructing comprise:

(b-1) constructing a remaining document common co-occurrence matrix for the set of the remaining documents;

(b-2) obtaining a document commonality of each remaining document to the set of the remaining documents based on a product sum between every component of the co-occurrence matrix of each remaining document and the corresponding component of the remaining document common co-occurrence matrix;

(b-3) extracting, as the seed document, the document having the highest document commonality to the set of the remaining documents; and

(b-4) constructing the initial cluster by including the seed document and neighbor documents similar to the seed document;

(c) making documents, which have the document commonality to the current cluster higher than a threshold, belong temporarily to the current cluster; wherein said making comprising:

(c-1) constructing a current cluster common co-occurrence matrix for the current cluster and a current cluster document frequency matrix of the current cluster based on occurrence frequencies of terms or term pairs appearing in the documents of the current cluster;

(c-2) obtaining a distinctiveness value of each term and each term pair for the current cluster by comparing the input document frequency matrix with the current cluster document frequency matrix;

(c-3) obtaining weights of each term and each term pair from their distinctiveness values;

(c-4) obtaining a document commonality to the current cluster for each document in the input document set based on a product sum between every component of the co-occurrence matrix of the input document and the corresponding component of the current cluster common co-occurrence matrix while applying the respective weights to said components; and

(c-5) making documents having the document commonality to the current cluster higher than the threshold belong temporarily to the current cluster;

(d) repeating step (c) until the number of documents temporarily belonging to the current cluster does not increase;

(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and

(f) deciding, on the basis of the document commonality of each document to each cluster, a cluster to which each document belongs and outputting said cluster.

The amended claim language finds support in the application as filed as follows:

step (a)	blocks 21, 22, 23 of figure 2A and paragraphs 0050 (co-occurrence matrix for each input document), and 0054 (input document frequency matrix for the set of input documents)
step (b)	block 24 of figure 2A and the corresponding text in the specification
step (b-1)	Equation in paragraph 0053
step (b-2)	Equation in paragraph 0059
step (b-3)	Paragraph 0062
step (b-4)	Paragraph 0062
step (c)	block 25 of figure 2B and the corresponding text in the specification
step (c-1)	blocks 31, 32 of figure 3 and equation in paragraph 0053, as well as paragraph 0054 (input document frequency matrix for the set of input documents)
step (c-2)	block 33 of figure 3 and equation in paragraph 0065
step (c-3)	block 33 of figure 3 and equation in paragraph 0067
step (c-4)	block 34 of figure 3 and equation in paragraph 0069
step (c-5)	block 35 of figure 3 and the corresponding text in the specification
step (d)	block 25 of figure 2B and the corresponding text in the specification
step (e)	blocks 26, 27 of figure 2B and the corresponding text in the specification
step (f)	block 29 of figure 2B and the corresponding text in the specification

Amended independent claim 1 is patentable over the art at least because none of the applied references teach or suggest the steps (b-1) and (c-1) which find support in the equation in paragraph 0053. Therefore, the art also fails to teach or suggest the steps (b-2) and (c-4), which find support in the equations in paragraphs 0059 and 0069, because the steps (b-2) and (c-4) use the co-occurrence matrix obtained in steps (b-1) and (c-1), respectively.

In addition, the iteration of steps (c-1) through (c-5), especially (c-1) through (c-4), results in advantages disclosed in paragraphs 0032 and 0085. Namely, document commonality has a large value for the document that is close to the topic to thereby improve clustering accuracy. The correct

number of clusters can be extracted from an input document set, and each document can be assigned to a cluster with high accuracy. Therefore, the efficiency of information acquisition by the user can be significantly improved.

Accordingly, amended claim 1 is patentable over the art.

Independent claim 29 has been amended in substantially the same manner and should be considered patentable for at least the reasons presented with respect to claim 1.

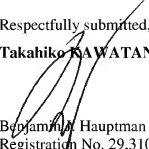
The dependent claims, including any new claim(s), are considered patentable at least for the reason(s) advanced with respect to the respective independent claim(s).

Each of the Examiner's rejections has been traversed/overcome. Accordingly, Applicants respectfully submit that all claims are now in condition for allowance. Early and favorable indication of allowance is courteously solicited.

The Examiner is invited to telephone the undersigned, Applicant's attorney of record, to facilitate advancement of the present application.

To the extent necessary, a petition for an extension of time under 37 C.F.R. 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account 08-2025 and please credit any excess fees to such deposit account.

Respectfully submitted,  
**Takahiko KAWATANI**



Benjamin Hauptman  
Registration No. 29,310

**HEWLETT-PACKARD COMPANY**

Intellectual Property Administration  
P.O. Box 272400  
Fort Collins, CO 80527-2400  
Telephone: 703-684-1111  
Facsimile: 970-898-0640  
Date: July 15, 2008  
BJH:KL/tal